



Epigenome-Wide Assessment of DNA Methylation in the Placenta and Arsenic Exposure in the New Hampshire Birth Cohort Study (USA)

**Benjamin B. Green, Margaret R. Karagas, Tracy Punshon,
Brian P. Jackson, David J. Robbins, E. Andres Houseman,
and Carmen J. Marsit**

<http://dx.doi.org/10.1289/ehp.1510437>

Received: 7 July 2015

Accepted: 18 December 2015

Advance Publication: 15 January 2016

Note to readers with disabilities: *EHP* will provide a [508-conformant](#) version of this article upon final publication. If you require a 508-conformant version before then, please contact ehp508@niehs.nih.gov. Our staff will work with you to assess and meet your accessibility needs within 3 working days.



National Institute of
Environmental Health Sciences

Epigenome-Wide Assessment of DNA Methylation in the Placenta and Arsenic Exposure in the New Hampshire Birth Cohort Study (USA)

Benjamin B. Green^{1,2,3}, Margaret R. Karagas^{1,3}, Tracy Punshon⁴, Brian P. Jackson⁵,
David J. Robbins⁶, E. Andres Houseman⁷, and Carmen J. Marsit^{1,2,3}

¹Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire, USA; ²Department of Pharmacology and Toxicology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA; ³Children's Environmental Health and Disease Prevention Research Center at Dartmouth Geisel School of Medicine, Lebanon, New Hampshire, USA; ⁴Department of Biological Sciences, Dartmouth College, Hanover, New Hampshire, USA; ⁵Department of Earth Sciences, Dartmouth College, Hanover, New Hampshire, USA; ⁶Molecular Oncology Program, The DeWitt Daughtry Family Department of Surgery, University of Miami Miller School of Medicine, Miami, Florida, USA; ⁷School of Biological and Population Health Sciences, College of Public Health and Human Sciences, Oregon State University, Corvallis, Oregon, USA

Address correspondence to Carmen J. Marsit, Department of Epidemiology, Geisel School of Medicine at Dartmouth, 7650 Remsen, Hanover, NH 03755 USA. Telephone: (603)-650-1825. E-mail: Carmen.J.Marsit@dartmouth.edu

Running title: Arsenic exposure and placental DNA methylation

Acknowledgments: The authors would like to acknowledge support for this study from NIH-NIEHS P01 ES022832, US EPA grant RD83544201, NIH-NIGMS P20GM104416, and NCI R25CA134286. Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the US EPA. Further, the US EPA does not endorse the purchase of any commercial products or services mentioned in the presentation.

Competing financial interests: The authors declare they have no actual or potential competing financial interests.

Abstract

Background: Arsenic is one of the most commonly encountered environmental toxicants, and research from model systems has suggested that one mode of its toxic activity may be through alterations in DNA methylation. *In utero* exposure to arsenic can impact fetal, newborn, and infant health resulting in a range of phenotypic outcomes.

Objectives: This study examined variation in placental DNA methylation and its relationship to arsenic exposure in 343 individuals enrolled in the New Hampshire Birth Cohort Study.

Methods: Linear regression models using a reference-free correction to account for cellular composition were employed to determine CpG loci affected by arsenic levels.

Results: Total arsenic measured in maternal urine during the second trimester was not associated with methylation in the placenta, while arsenic levels quantified through maternal toenail collected at birth were associated with methylation at a single CpG loci ($p = 4.1 \times 10^{-8}$). Placenta arsenic levels were associated with 163 differentially methylated loci ($FDR < 0.05$), with 11 probes within the *LYRM2* gene reaching genome-wide significance ($p < 10^{-8}$). Measurement of *LYRM2* mRNA levels indicated that methylation was weakly to moderately correlated with expression ($r=0.15$, $p < 0.06$). In addition, we identified pathways suggesting changes in placental cell subpopulation proportions associated with arsenic exposure.

Conclusions: These data demonstrate the potential for arsenic, even at levels commonly experienced in a US population, to have effects on the DNA methylation status of specific genes in the placenta and thus supports a potentially novel mechanism for arsenic to impact long-term children's health.

Introduction

One hundred million individuals worldwide, 13 million of which reside within the United States, are exposed to arsenic in drinking water at levels above the World Health Organization and US Environmental Protection Agency recommended limit of 10 µg/L (EPA 2009). Arsenic exposure often originates from the use of unregulated wells for drinking water as well as from dietary sources, with levels of exposure varying widely depending on the population and region studied (EFSA 2009; Gilbert-Diamond et al. 2011; WHO 2011). Arsenic exposure during pregnancy is of special concern to fetal health as it easily crosses the placenta, exposing the infant to levels comparable to those of the mother (Concha et al. 1998). Fetal exposure to arsenic has previously been linked to negative birth outcomes including a general increase in infant mortality (Rahman et al. 2007; Rahman et al. 2010), low birth weight (Germann et al. 2013; Laine et al. 2014), poor infant neurodevelopment (Hamadani et al. 2010; Hamadani et al. 2011), as well as increased infant infection rates and alterations in their immune profile (Farzan et al. 2013; Nadeau et al. 2014).

While several studies have investigated the associations between arsenic exposure and epigenetic alterations in both adult blood (Argos et al. 2015; Bailey et al. 2013; Liu et al. 2014; Seow et al. 2014) and cord blood (Kile et al. 2014; Koestler et al. 2013), there is a gap in our understanding of the role of maternal arsenic exposure as it relates to fetal DNA methylation on a genome-wide scale within the placenta. In existing genome-scale studies, DNA examined from umbilical cord blood samples indicated differential methylation in specific regions across the genome, with variation enriched at loci within CpG islands (Koestler et al. 2013). In addition, employing a statistical model to estimate

leukocyte proportions from DNA methylation array data (Houseman et al. 2012), an increase in the proportion of CD8⁺ T-cells was described as a function of arsenic exposure (Koestler et al. 2013), a finding that was later replicated in a Bangladeshi population exposed to a wider range of arsenic (Kile et al. 2014).

The placenta may be an ideal candidate for investigations of fetal programming and environmental impacts on long-term health, because it mediates nutrient and waste exchange, regulates interaction with the maternal immune system, acts as a neuroendocrine organ, and regulates fetal exposure to exogenous compounds present in the maternal circulation, making it a master regulator for the fetal environment (Maccani and Marsit 2009; Novakovic and Saffery 2012). Recent studies have linked variation in DNA methylation in the placenta with a number of environmental exposures as well as newborn health outcomes including growth and neurobehavioral functioning (Appleton et al. 2015; Lesseur et al. 2014; Maccani et al. 2015; Simpkin et al. 2015). We hypothesized that maternal arsenic exposure can impact the function of the placenta through variation in the placental DNA methylation profile. To assess the relationship between arsenic and the placental epigenome, we examined samples originating from the New Hampshire Birth Cohort Study (NHBCS), where individuals have varying levels of arsenic exposure from use of unregulated private wells for drinking water as well from dietary exposure (Farzan et al. 2013). Genome-wide DNA methylation was measured in placenta samples and the relationship between methylation and various biomarkers of maternal and fetal arsenic exposure, including maternal urine and toenail samples as well placenta levels, was assessed. Further, we investigated how arsenic associated variable DNA methylation was correlated with gene expression at identified regions. To our knowledge, this study

presents the first genome-wide analysis of placental DNA methylation as it relates to arsenic exposure *in utero*.

Materials and Methods

Study Population

The study population consisted of 343 mother-infant pairs who were participants in the New Hampshire Birth Cohort Study, an ongoing prospective study initiated in 2009. Participants chosen for these analyses included all mothers recruited into the cohort from February of 2012 to September of 2013. All subjects provided written informed consent in accordance with the requirements of the Institutional Review Board (IRB) of Dartmouth College. Briefly, mothers enrolled in the study were between 18-45 years of age, reporting use of a private, unregulated well at their home, pregnant with a singleton infant and not planning to move.

Sample Collection

Maternal urine samples were collected at approximately 24-28 weeks gestation following protocols described elsewhere (Gilbert-Diamond et al. 2011). Maternal toenail samples were obtained following delivery, as previously described (Davis et al. 2014). Following delivery, the placenta was biopsied adjacent to the cord insertion (to minimize heterogeneity), removing any maternal decidua, placed immediately in RNeasy (Life Technologies, Carlsbad, CA) and frozen at -80°C within 24 hours. DNA and RNA were extracted by using the RNA/DNA extraction kit (Norgen Biotek, Thorold, ON) subsequently quantified using the Qubit Fluorometer (Life Technologies) and stored at -80 °C.

Arsenic Quantification

Arsenic levels were quantified within maternal urine (collected between 24-28 weeks gestation), maternal post-partum toenail samples and placenta from 342 individuals with downstream analysis conducted on samples with values above the detectable limit.

Arsenic in all three tissues was quantified using inductively coupled plasma dynamic reaction cell-mass spectrometry. The protocol has been described previously (Davis et al. 2014; Punshon et al. 2015). A total of 285 samples had detectable placental arsenic data, and for these the detection limit for arsenic in placenta was 0.0148 ng/g. Total urinary arsenic was calculated as the sum of arsenate, arsenite, monomethylarsonic acid (MMA), and dimethylarsinic acid (DMA). Detection limits for the four individual arsenic species ranged from 0.10 to 0.15 mg/l and 144, 138, 59, and 3 of 271 samples with available data were below detection limit for arsenate, arsenite, MMA, and DMA, respectively. For samples with one of these individual species below the detection limit, we assigned values equal to the detection limit divided by the square root of two (Helsel 1990). Arsenic was detected in 257 of the maternal toenail samples. We further performed Pearson correlation tests on \log_{10} -transformed arsenic values from the three tissues.

DNA Methylation Quantification and Normalization

Isolated DNA underwent bisulfite modification using the EZ Methylation kit (Zymo Research, Irvine, CA). Samples of bisulfite modified DNA were then randomized across several plates and assessed for epigenome-wide DNA methylation levels using the Infinium HumanMethylation450 (450K) Bead Chip (Illumina, San Diego, CA) profiling methylation status for approximately 486,000 CpG loci. The microarrays were processed at the Biomedical Genomics Center at the University of Minnesota (Minneapolis, MN).

Data was assembled using the BeadStudio methylation software package (Illumina) and then processed using the ‘minfi’ package in R (R Core Team 2013). For our analysis, we removed all probes that were present on either the X or Y chromosome, have previously been identified as cross-hybridizing with other genomic locations (Chen et al. 2013), or contained a single nucleotide polymorphism. In addition, probes identified as having detection P values >0.01 in at least one sample were removed. This resulted in 344,348 autosomal probes from 343 unique samples being included within our analysis.

Data were normalized using the functional normalization (funNorm) protocol within the ‘minfi’ package as specified by the software authors (Guide available at: <http://bioconductor.org/packages/release/bioc/html/minfi.html>). A principle component analysis (PCA) was then used to determine whether technical aspects of our design (bisulfite conversion plate or microarray slide) might be influencing underlying variation within the data set. We examined the two principle components accounting for the largest degree of variation and found that plate was associated with a significant portion of the variation. In order to correct this, we adjusted our data for this effect using the ‘ComBat’ method (Johnson et al. 2007), and again used PCA to assure this bias was corrected. Raw and processed data has been provided via the NCBI Gene Expression Omnibus (GEO) accession number GSE71678.

Cell Mixture Deconvolution and Analysis

Using normalized and ComBat adjusted methylation data, we proceeded to investigate the relationship between placental methylation and measured levels of arsenic exposure. Of the 343 samples within our study, 285, 257, and 271 had available placenta, toenail, and total urinary arsenic measurements, respectively. These subsets were used to

determine the relationship between arsenic and methylation. The relationships between methylation and each measure of arsenic were evaluated via epigenome-wide association studies (EWAS) using the R package RefFreeEWAS which fits a series of linear models and subsequently accounts for cell proportion variability within tissue samples without a reference set (Houseman et al. 2014). While the reference free approach does not require *a priori* knowledge of cell composition within a tissue to accomplish the adjustment, placenta tissue is likely a combination of variously differentiated trophoblasts, stromal cells including fibroblasts, and mesenchymal cells including Hofbauer and immune cells (Wang and Zhao 2010). Using DNA methylation as the dependent variable, we performed an EWAS with arsenic levels in placental, maternal urine, or maternal toenail as the explanatory variable, while also including maternal age, gestational age, and infant sex as potential confounding covariates within the model. Maternal and gestational ages were modeled as continuous variables. A separate EWAS was performed for each arsenic measurement from the three tissues. Genome-wide significance based on a Bonferroni threshold was considered at $p < 1.0 \times 10^{-7}$, though we report all loci which passed an FDR correction of $q < 0.05$. All p-values reported are the product of the multiple comparisons correction.

The RefFreeEWAS method produces an effect estimate for each of the CpG loci investigated and returns results for both a covariate-only adjusted (β) and a covariate and cell-mixture adjusted (β^*) model. Those loci exhibiting the large differences between the β and β^* estimates would represent those most confounded by differences in cell subpopulation proportions. By linking the genes associated with these highly confounded CpG sites to gene ontology (GO) terms, we can make an indirect assessment of the

representation of these genes within specific functions or pathways, and infer then the types of cellular functions and thus types of cells, most likely demonstrating altered subpopulations by arsenic exposure. To decide which loci to include, we calculated the difference (δ) between β^* and β for each loci, and used the resulting value and standard error to derive a p-value for δ , with a null-hypothesis of $\beta^* = \beta$. We considered all loci with $p < 0.10$ to have undergone significant cell mixture corrections while including placental arsenic levels within the model. A Cochran-Mantel-Haenszel test was then employed to determine associations between these loci and gene ontology (GO) term pathways while accounting for the loci's relation to a CpG island as well as position relative to a gene. This would then allow us to indirectly determine how exposure to arsenic *in utero* may be altering cellular composition within the placenta based upon pathways over-represented amongst these loci.

Validation of previously identified loci associated with arsenic exposure

Several recent papers have investigated the role of arsenic exposure on an epigenome-wide scale. While none of these reports have examined placenta-specific DNA methylation, the use of an identical or similar platform does allow for some comparisons to be made in the findings. These studies investigated the effects of arsenic on DNA methylation as measured in both adult (Argos et al. 2015; Liu et al. 2014; Seow et al. 2014) and cord (Kile et al. 2014; Koestler et al. 2013; Rojas et al. 2015) blood or in a single study, urothelial carcinomas (Yang et al. 2014). We examined these loci by linear regression model using methylation as the explained variable and placental arsenic as the explanatory variable.

LYRM2 Gene Expression and Analysis

The expression status of *LYRM2* was examined, as the methylation status of 11 neighboring CpG loci within the promoter of this gene were found to be genome-wide significantly associated with arsenic exposure. Placental RNA samples from a subset of 96 individuals, selected from the highest and lowest tertile of average methylation for the gene *LYRM2*, were used for gene expression analysis. Due to sample failure, we were able to collect expression data from 93 individuals. For each sample, cDNA as well as a non-template control were produced using the iScript cDNA master mix (Bio-Rad, Hercules, CA) as specified within manufacturer's instructions and stored at -4°C. mRNA levels were quantified using the iQ SYBER Green Supermix (Bio-Rad) on the CFX Connect Real-Time PCR Detection System (Bio-Rad) per manufacturer's instructions. Primers specific to *LYRM2* were used from the TaqMan Gene Expression Assays kit (Life Technologies, Cat #4351372) following recommended PCR conditions. Expression levels were normalized against the housekeeping gene succinate dehydrogenase (*SDHA*) using TaqMan Gene Expression Assay primers (Life Technologies). We selected *SDHA* for our housekeeping gene as it has previously been shown to be the most effective for use in placental gene expression studies due to stable expression across samples (Meller et al. 2005).

Pearson correlation coefficients were calculated to determine the relationship between DNA methylation within the *LYRM2* gene and expression levels. *LYRM2* methylation was input as the average β value for the 11 probes that were identified by our previous modeling. All statistical analyses were conducted within the R statistical package, version 3.1.0 (Vienna, Austria, www.r-project.org).

Results

The characteristics of the 343 mother-infant pairs examined in this analysis are presented in **Table 1**. Arsenic levels in all media had a broad range across individuals with a general skew towards the right. Mothers were on average approximately 31.5 years of age at the time of birth, with a BMI of 25.6 calculated from pre-pregnancy weight. There was an approximately even distribution of male and female infants within the study population with an average of 39.4 weeks gestation and 3.4kg birth weight. We also determined the relationship between the measures of arsenic for the three tissues as calculated from the 176 individuals with values for all three biological samples. There was no correlation between the measures of arsenic within the placenta and maternal toenail samples ($r = 0.03$), or between placenta and urine ($r = 0.03$), and only a weak correlation between maternal toenail samples and urine ($r = 0.19$).

Focusing on our study, 163 CpG dinucleotides demonstrated significant associations ($q < 0.05$) with placental arsenic concentration (**Excel File Table S1**). Of those, thirteen attained genome-wide significance following multiple comparisons correction with the more stringent Bonferroni correction method (**Figure 1A**). Among these probes one (*cg08640609*) tracked to calmodulin binding transcription activator-1 (*CAMTA1*), one (*cg01777586*) to coiled-coil domain containing-57 (*CCDC57*), and eleven to LYR-motif containing 2 (*LYRM2*) (**Table 2**). Only one site (*cg01021483*), within the 3'UTR of mini chromosome maintenance complex component 5 (*MCM5*), was genome-wide significantly ($q < 0.05$, $P = 1.76 \times 10^{-8}$) associated with toenail arsenic (**Figure 1B**). No loci demonstrated significant associations to total urinary arsenic (**Figure 1C**). Estimate effects and corresponding p-values were plotted for all three

measures of arsenic exposure (**Figure 2**) and the majority of loci with methylation demonstrating significant associations with placental arsenic were hypomethylated with increasing arsenic exposure. None of the loci demonstrating a relationship between placental arsenic and DNA methylation showed any association when urinary or toenail arsenic were used as the explanatory variable after correcting for multiple comparisons (**Excel File Table S2**).

The eleven probes associated with *LYRM2* were in the promoter region within approximately 600bp of one another. These probes all showed a negative relationship with arsenic, demonstrating a decreased level of methylation as arsenic exposure increased. The methylation β -values displayed a strong positive correlation between each of the eleven sites, allowing downstream analysis to focus on mean methylation levels for these sites in order to reduce the number of comparisons being made (**Figure 3**). In a subsample of 93 individuals from the highest and lowest tertile of methylation for *LYRM2* loci, we observed a moderate negative correlation ($r = -0.15$) between *LYRM2* expression within the placenta and mean promoter methylation that approached ($p = 0.058$) statistical significance (**Figure 4**).

Due to the use of the RefFreeEWAS methodology, in addition to examining CpG loci with variable methylation most associated with placental arsenic exposure independent of cellular mixture, we can also identify those loci whose effect estimates are changed the most when cell mixture is controlled. Of the 344,348 loci analyzed from the methylation array, 10,596 sites showed substantial (Bonferroni corrected, $p < 0.10$) correction based upon the δ estimate, representing the difference in effect estimates of placental arsenic on methylation with and without cell mixture control. 319 GO-terms

demonstrated significant over-representation (Bonferroni corrected, $p < 0.05$) amongst loci that were most effected by cellular subpopulation correction. The top 20 GO-terms over-represented amongst the genes associated with these loci are provided in **Table 3**.

In the present study, we evaluated previously reported loci from other arsenic based epigenome-wide studies (Argos et al. 2015; Kile et al. 2014; Liu et al. 2014; Rojas et al. 2015; Seow et al. 2014; Yang et al. 2014) using a lookup approach to determine whether they presented a similar sensitivity in the placenta. In total, 3,393 CpG loci highlighted from these publications were also investigated within our array, though it is important to note the majority of these loci were identified within a single study (3,316 from Rojas, et al. 2015). Associations were identified for 293 of these CpG sites ($p < 0.05$, no adjustment for multiple comparisons) within our data set. Again, the majority of these were from the single Rojas et al. study, though 6 of the loci identified were from the other studies examined (**Excel File Table S3**).

Discussion

This study examined the association between DNA methylation within the placenta and arsenic exposure measured via three surrogate tissues. Arsenic measurements from maternal urine (median = 3.76 $\mu\text{g/L}$) were comparable to prior subsets examined from within this cohort (median = 4.1 $\mu\text{g/L}$) (Koestler et al. 2013), though are lower than values obtained from a highly exposed Bangladeshi population (median = 12 $\mu\text{g/L}$) (Kile et al. 2014) as well as a representative sampling of the National Health and Nutrition Examination Survey participants (median = 4.2 $\mu\text{g/L}$) (Kuo et al. 2015). In addition, arsenic levels from the toenail, a measurement of long-

term exposure was comparable, though slightly lower (IQR = 0.026 $\mu\text{g/g}$ to 0.073 $\mu\text{g/g}$) than values from the Coronary Artery Risk Development in Young Adults Study (IQR = <0.0649 $\mu\text{g/g}$ to >0.1442 $\mu\text{g/g}$). While there are no studies with comparable measurements of arsenic within the placenta, the arsenic values for urine and toenail samples from this cohort are within the range of other reported values, suggesting that the values we measured within the placenta would also be generalizable to exposure within other US based cohorts. A prior study in this cohort examining a much larger sample (n=766) observed a significant relationship between placental arsenic levels and maternal urine or toenail (r_s = 0.11 and 0.12, respectively) (Punshon et al. 2015). In the present study, we saw only weak if any correlations between the biomarkers, likely due to the reduced sample size examined. Urinary arsenic may be considered a measurement of acute, recent exposure to arsenic (within the 24-28 week gestational timeframe at which samples were taken), in contrast to toenail samples, which give a view of more long-term exposures and storage within this tissue. The placenta likely represents a measurement of arsenic exposure and deposition within the window of the pregnancy. Thus, our disparate findings of variable DNA methylation by each of these biomarkers are not unexpected.

Using placental tissue from 343 mother-infant pairs, we identified that variation in DNA methylation of 163 CpG loci was statistically associated with arsenic levels at a 5% FDR. Among these 163 sites, 15 within the promoter region of *LYRM2* were linked to placental arsenic exposure burden, with 11 of these sites maintaining significance after the more conservative Bonferroni multiple tests correction. Further, we provide evidence that the loci identified within the *LYRM2* promoter region may influence gene transcription, consistent with the canonical role of promoter DNA methylation on gene

transcription potential. Samples with the highest levels of methylation showed a trend toward muted expression levels, while samples with low methylation displayed a range of expression, though contained a high proportion of individuals with the highest levels of expression. The variability in the relationship between methylation and expression observed in this data, particularly at the lowest extents of DNA methylation, is expected, as the absence of methylation does not guarantee high expression levels, but instead endows a gene with the *potential* for expression. In addition, it is important to note that while DNA methylation is a major contributor to regulation of gene expression, there is not always a guaranteed relationship between these two biological measurements. Because of this, it is important to examine the relationship between identified differentially methylated regions and the expression of the associated gene, and highlights the strength of the nearly significant relationship identified within this study.

LYRM2 is a member of the LYR-motif containing family of genes. While *LYRM2* currently lacks a known function within the placenta or other tissues, several members of this gene family have been ascribed an important role in iron-sulfur cluster (ISC) protein production (Angerer 2015; Maio et al. 2014). Specifically, researchers using a high-throughput yeast two-hybrid approach described an interaction between the C-terminus of the ISC chaperone protein, HSC20, and proteins bearing the LYR motif (Maio et al. 2014), suggesting that this motif targets proteins for ISC delivery. This result built upon previous reports linking *LYRM4* (also known as *ISD11*) as a binding partner to *NFS1* cysteine desulfurase (*NFS1*), a sulfur donor in ISC biogenesis (Shi et al. 2009). Deficits in ISC-assembly manifest in several different forms in human health and disease including myopathy and neurodegenerative disease (Rouault and Tong 2008; Rouault

2012). In addition, several other members of the LYR superfamily are associated with the electron transport chain, specifically in the assembly of complex subunits I, II, III, and V (Angerer 2015). While it is currently speculative to insert LYRM2 into these processes, the proportion of LYR-motif containing genes associated with both the electron transport chain and ISC biogenesis would strongly suggest a role for LYRM2. Data from these prior publications would also suggest that proteins containing the LYR motif are targets of ISC delivery (Angerer 2015; Maio et al. 2014). This is intriguing, as iron sulfur cluster containing proteins can act as a reductase involved in arsenic metabolism in anaerobic organisms (Saltikov and Newman 2003). While no studies to our knowledge have implicated ISC proteins in eukaryotic arsenic metabolism, these findings may still potentially link arsenic to alterations in a gene with functional implications in the response to the exposure.

While the majority of loci identified resided within the *LYRM2* promoter, two CpG sites linked to separate genes were also found to be associated with placenta arsenic exposure. One site is located within the promoter region of *CAMTA1*, a member of the highly conserved family of transcription factors associated with calcium signaling through a direct binding to calmodulin (Bouche et al. 2002). Micro-rearrangements of this region have been implicated in congenital ataxia (Thevenon et al. 2012) while deletions have been associated with the progression of neuroblastomas (Katoh and Katoh 2003). The second site is located within the *CCDC57* promoter. While the function of this gene is unclear, it has been identified to play a role in the interactome of nutritional metabolomics (Shin et al. 2014). In addition, overexpression of MCM5, identified here as associated with maternal toenail arsenic levels, has previously been identified in a variety

of human cancers (Gakiopoulou et al. 2007; Guida et al. 2005; Liu et al. 2007). The biological implications of these genes are unclear as they relate to arsenic toxicity and arsenic related birth outcomes and further work is required to elucidate mechanistic involvement of these genes, and to assess if the methylation identified has any functional impact on these genes.

In addition to identifying regions impacted by arsenic exposure, we also investigated the potential role of arsenic on alterations to placental cellular composition by examining which loci demonstrated effect estimates undergoing substantial change when cellular composition was included in the model. Such analyses of changes in blood cell composition based on estimates derived from DNA methylation data have been performed in studies of cancer as well as maternal arsenic and mercury exposures and infant umbilical cord blood (Cardenas et al. 2015; Kile et al. 2014; Koestler et al. 2013). We could not accomplish this directly, as we could not directly examine cellular subpopulations based on histologic or morphologic criteria, and there is currently no reference set of DNA methylation data for placenta cell subsets, such as those that exist for blood samples (Reinius et al. 2012). Instead, we used a computational approach to examine if the loci with the greatest change in effect estimate due to a correction for cell mixture were over-represented amongst cellular pathways that could inform on specific cellular functions indicative of specific cell population, similar to analyses recently undertaken (Houseman et al. 2015). Our analysis noted (**Table 3**) that loci affected by arsenic-associated cellular composition changes were significantly associated with morphogenic and developmental processes such as, anterior/posterior pattern specification, embryonic skeletal system morphogenesis, regulation of neuron

differentiation, and forebrain neuron development, among others, suggesting potential important effects on placental development.

Previous investigations have also focused on the role of arsenic exposure in altering epigenetic signatures from a number of tissues, including both adult (Argos et al. 2015; Liu et al. 2014; Seow et al. 2014) and cord blood (Kile et al. 2014; Koestler et al. 2013; Rojas et al. 2015), and in a single study, urothelial carcinomas (Yang et al. 2014). We compared the findings from those studies with our data and found an overlap in 293 of 3,393 previously identified differentially methylated loci. It is also important to note that we did not perform a correction for multiple testing to the resulting p-value since this was a targeted lookup of association rather than a major exploratory analysis. Should these sites have been investigated without the *a priori* knowledge of the previous associations, the significance would not have held up following the epigenome-wide correction for multiple tests. Nonetheless, the associations between placental arsenic and the previously identified sites do lend a modest level of validation to their results. In addition, the significance seen in our study for these loci is of note in that none of these prior studies utilized placental tissue, suggesting there may be, to a certain degree, conservation across tissue types of the role of arsenic on methylation levels of a small proportion of genomic regions.

This study has several potential limitations. Our samples were collected from placentas at term, reducing the ability to infer how arsenic associated methylation variability plays a role throughout development. Although this is one of the largest studies, to date, linking maternal exposures to infant epigenetic features, we are still limited by our sample size in our ability to observe robust associations with potentially

greater variability or small effect sizes. It will also be important, as additional populations begin to assess placenta DNA methylation, to replicate these findings in additional cohorts. With this in mind however, it is important to note the many strengths of the current study. Our study offers perspective on a relatively healthy population exposed to arsenic levels in line with that of the general US population. In addition, the measurement of arsenic levels within the placenta gives our study a direct assessment of likely arsenic levels present during gestation, rather than through the use of a proxy tissue.

Conclusions

Arsenic exposure *in utero* has been implicated in a myriad of negative birth outcomes in infants, yet the mechanisms by which these changes occur are poorly understood. Here we find that arsenic exposure as measured within the placenta is associated with variation in placental DNA methylation, with a particularly strong association found with methylation of *LYRM2*. Data to replicate this study were unavailable. As we increase our understanding of the role of arsenic in infant health, should the results presented here be confirmed by other studies these data may highlight the role of the placenta in regulating arsenic exposure during pregnancy.

References

- Angerer H. 2015. Eukaryotic lrr proteins interact with mitochondrial protein complexes. *Biology* 4:133-150.
- Appleton AA, Lester BM, Armstrong DA, Lesseur C, Marsit CJ. 2015. Examining the joint contribution of placental nr3c1 and hsd11b2 methylation for infant neurobehavior. *Psychoneuroendocrinology* 52:32-42.
- Argos M, Chen L, Jasmine F, Tong L, Pierce BL, Roy S, et al. 2015. Gene-specific differential DNA methylation and chronic arsenic exposure in an epigenome-wide association study of adults in Bangladesh. *Environ Health Perspect* 123:64-71.
- Bailey KA, Wu MC, Ward WO, Smeester L, Rager JE, García-Vargas G, et al. 2013. Arsenic and the epigenome: Interindividual differences in arsenic metabolism related to distinct patterns of DNA methylation. *Journal of Biochemical and Molecular Toxicology* 27:106-115.
- Bouche N, Scharlat A, Snedden W, Bouchez D, Fromm H. 2002. A novel family of calmodulin-binding transcription activators in multicellular organisms. *The Journal of biological chemistry* 277:21851-21861.
- Cardenas A, Koestler DC, Houseman EA, Jackson BP, Kile ML, Karagas MR, et al. 2015. Differential DNA methylation in umbilical cord blood of infants exposed to mercury and arsenic in utero. *Epigenetics*:1-8.
- Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. 2013. Discovery of cross-reactive probes and polymorphic CpGs in the illumina Infinium humanmethylation450 microarray. *Epigenetics* 8:203-209.
- Concha G, Vogler G, Lezcano D, Nermell B, Vahter M. 1998. Exposure to inorganic arsenic metabolites during early human development. *Toxicological sciences : an official journal of the Society of Toxicology* 44:185-190.
- Davis MA, Li Z, Gilbert-Diamond D, Mackenzie TA, Cottingham KL, Jackson BP, et al. 2014. Infant toenails as a biomarker of in utero arsenic exposure. *Journal of exposure science & environmental epidemiology* 24:467-473.
- EFSA. 2009. EFSA panel on contaminants in the food chain, scientific opinion on arsenic in food.EFSA.

- EPA US. 2009. Fact sheet: Drinking water standard for arsenic. Available:
<http://www.epa.gov/dwreginfo/chemical-contaminant-rules>
- Farzan SF, Korrick S, Li Z, Enelow R, Gandolfi AJ, Madan J, et al. 2013. In utero arsenic exposure and infant infection in a united states cohort: A prospective study. *Environmental Research* 126:24-30.
- Gakiopoulou H, Korkolopoulou P, Levidou G, Thymara I, Saetta A, Piperi C, et al. 2007. Minichromosome maintenance proteins 2 and 5 in non-benign epithelial ovarian tumours: Relationship with cell cycle regulators and prognostic implications. *British journal of cancer* 97:1124-1134.
- Gelmann ER, Gurzau E, Gurzau A, Goessler W, Kunrath J, Yeckel CW, et al. 2013. A pilot study: The importance of inter-individual differences in inorganic arsenic metabolism for birth weight outcome. *Environmental toxicology and pharmacology* 36:1266-1275.
- Gene Ontology Consortium. 2015. Gene Ontology Consortium Database. Available:
geneontology.org. [accessed June 2015].
- Gilbert-Diamond D, Cottingham KL, Gruber JF, Punshon T, Sayarath V, Gandolfi AJ, et al. 2011. Rice consumption contributes to arsenic exposure in US women. *Proc Natl Acad Sci U S A* 108:20656-20660.
- Guida T, Salvatore G, Faviana P, Giannini R, Garcia-Rostan G, Provitera L, et al. 2005. Mitogenic effects of the up-regulation of minichromosome maintenance proteins in anaplastic thyroid carcinoma. *J Clin Endocrinol Metab* 90:4703-4709.
- Hamadani JD, Grantham-McGregor SM, Tofail F, Nermell B, Fangstrom B, Huda SN, et al. 2010. Pre- and postnatal arsenic exposure and child development at 18 months of age: A cohort study in rural Bangladesh. *Int J Epidemiol* 39:1206-1216.
- Hamadani JD, Tofail F, Nermell B, Gardner R, Shiraji S, Bottai M, et al. 2011. Critical windows of exposure for arsenic-associated impairment of cognitive function in pre-school girls and boys: A population-based cohort study. *Int J Epidemiol* 40:1593-1604.
- Helsel DR. 1990. Less than obvious: Statistical treatment of data below the detection limit. *Environmental Science and Technology* 24:1766-1774.

- Houseman E, Accomando W, Koestler D, Christensen B, Marsit C, Nelson H, et al. 2012. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* 13:86.
- Houseman EA, Molitor J, Marsit CJ. 2014. Reference-free cell mixture adjustments in analysis of DNA methylation data. *Bioinformatics* 30:1431-1439.
- Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ. 2015. Cell-composition effects in the analysis of DNA methylation array data: A mathematical perspective. *BMC Bioinformatics* 16:95.
- Johnson WE, Li C, Rabinovic A. 2007. Adjusting batch effects in microarray expression data using empirical bayes methods. *Biostatistics (Oxford, England)* 8:118-127.
- Kato M, Kato M. 2003. Identification and characterization of flj10737 and camta1 genes on the commonly deleted region of neuroblastoma at human chromosome 1p36.31-p36.23. *International journal of oncology* 23:1219-1224.
- Kile ML, Houseman EA, Baccarelli AA, Quamruzzaman Q, Rahman M, Mostofa G, et al. 2014. Effect of prenatal arsenic exposure on DNA methylation and leukocyte subpopulations in cord blood. *Epigenetics* 9:774-782.
- Koestler D, Avissar-Whiting M, Houseman E, Karagas M, Marsit C. 2013. Differential DNA methylation in umbilical cord blood of infants exposed to low levels of arsenic in utero. *Environmental Health Perspectives* 121:971-977.
- Kuo CC, Weaver V, Fadrowski JJ, Lin YS, Guallar E, Navas-Acien A. 2015. Arsenic exposure, hyperuricemia, and gout in US adults. *Environment international* 76:32-40.
- Laine JE, Bailey KA, Rubio-Andrade M, Olshan AF, Smeester L, Drobna Z, et al. 2014. Maternal arsenic exposure, arsenic methylation efficiency, and birth outcomes in the biomarkers of exposure to arsenic (BEAR) pregnancy cohort in Mexico. *Environ Health Perspect.* 123:186-192.
- Lesseur C, Paquette AG, Marsit CJ. 2014. Epigenetic regulation of infant neurobehavioral outcomes. *Medical epigenetics* 2:71-79.
- Liu H, Takeuchi S, Moroi Y, Lin N, Urabe K, Kokuba H, et al. 2007. Expression of minichromosome maintenance 5 protein in proliferative and malignant skin diseases. *International journal of dermatology* 46:1171-1176.

- Liu X, Zheng Y, Zhang W, Zhang X, Liody-Jones DM, Baccarelli AA, et al. 2014. Blood methylomics in response to arsenic exposure in a low-exposed US population. *Journal of exposure science & environmental epidemiology* 24:145-149.
- Maccani JZ, Koestler DC, Lester B, Houseman EA, Armstrong DA, Kelsey KT, et al. 2015. Placental DNA methylation related to both infant toenail mercury and adverse neurobehavioral outcomes. *Environ Health Perspect.* 123:723-729.
- Maccani MA, Marsit CJ. 2009. Epigenetics in the placenta. *American Journal of Reproductive Immunology* 62:78-89.
- Maio N, Singh A, Uhrigshardt H, Saxena N, Tong W-H, Rouault Tracey A. 2014. Cochaperone binding to lyr motifs confers specificity of iron sulfur cluster delivery. *Cell Metabolism* 19:445-457.
- Meller M, Vadachkoria S, Luthy DA, Williams MA. 2005. Evaluation of housekeeping genes in placental comparative expression studies. *Placenta* 26:601-607.
- Nadeau KC, Li Z, Farzan S, Koestler D, Robbins D, Fei DL, et al. 2014. In utero arsenic exposure and fetal immune repertoire in a US pregnancy cohort. *Clinical immunology (Orlando, Fla)* 155:188-197.
- Novakovic B, Saffery R. 2012. The ever growing complexity of placental epigenetics – role in adverse pregnancy outcomes and fetal programming. *Placenta* 33:959-970.
- Punshon T, Davis MA, Marsit CJ, Theiler SK, Baker ER, Jackson BP, et al. 2015. Placental arsenic concentrations in relation to both maternal and infant biomarkers of exposure in a US cohort. *Journal of exposure science & environmental epidemiology.* 6:599-603.
- R Core Team. 2013. R: A language and environment for statistical computing. Vienna, Austria:R Foundation for Statistical Computing.
- Rahman A, Vahter M, Ekstrom EC, Rahman M, Golam Mustafa AH, Wahed MA, et al. 2007. Association of arsenic exposure during pregnancy with fetal loss and infant death: A cohort study in Bangladesh. *Am J Epidemiol* 165:1389-1396.
- Rahman A, Persson LA, Nermell B, El Arifeen S, Ekstrom EC, Smith AH, et al. 2010. Arsenic exposure and risk of spontaneous abortion, stillbirth, and infant mortality. *Epidemiology* 21:797-804.

- Reinius LE, Acevedo N, Joerink M, Pershagen G, Dahlen SE, Greco D, et al. 2012. Differential DNA methylation in purified human blood cells: Implications for cell lineage and studies on disease susceptibility. *PLoS One* 7:e41361.
- Rojas D, Rager JE, Smeester L, Bailey KA, Drobná Z, Rubio-Andrade M, et al. 2015. Prenatal arsenic exposure and the epigenome: Identifying sites of 5-methylcytosine alterations that predict functional changes in gene expression in newborn cord blood and subsequent birth outcomes. *Toxicological sciences : an official journal of the Society of Toxicology* 143:97-106.
- Rouault TA, Tong WH. 2008. Iron-sulfur cluster biogenesis and human disease. *Trends in genetics : TIG* 24:398-407.
- Rouault TA. 2012. Biogenesis of iron-sulfur clusters in mammalian cells: New insights and relevance to human disease. *Disease models & mechanisms* 5:155-164.
- Saltikov CW, Newman DK. 2003. Genetic identification of a respiratory arsenate reductase. *Proc Natl Acad Sci U S A* 100:10983-10988.
- Seow WJ, Kile ML, Baccarelli AA, Pan WC, Byun HM, Mostofa G, et al. 2014. Epigenome-wide DNA methylation changes with development of arsenic-induced skin lesions in Bangladesh: A case-control follow-up study. *Environ Mol Mutagen* 55:449-456.
- Shi Y, Ghosh MC, Tong WH, Rouault TA. 2009. Human ISD11 is essential for both iron-sulfur cluster assembly and maintenance of normal cellular iron homeostasis. *Hum Mol Genet* 18:3014-3025.
- Shin S-Y, Fauman EB, Petersen A-K, Krumsiek J, Santos R, Huang J, et al. 2014. An atlas of genetic influences on human blood metabolites. *Nat Genet* 46:543-550.
- Simpkin AJ, Suderman M, Gaunt TR, Lyttleton O, McArdle WL, Ring SM, et al. 2015. Longitudinal analysis of DNA methylation associated with birth weight and gestational age. *Hum Mol Genet* 24:3752-3763.
- Thevenon J, Lopez E, Keren B, Heron D, Mignot C, Altuzarra C, et al. 2012. Intragenic camt1 rearrangements cause non-progressive congenital ataxia with or without intellectual disability. *J Med Genet* 49:400-408.
- University of California Santa Cruz. 2015. Genome browser. Available: genome.ucsc.edu. [accessed June 2015].

- Wang Y, Zhao S. 2010. Vascular biology of the placenta. San Rafael, CA:Morgan & Claypool Life Sciences.
- WHO. 2011. Guidelines for drinking-water quality. Geneva:World Health Organization.
- Yang TY, Hsu LI, Chiu AW, Pu YS, Wang SH, Liao YT, et al. 2014. Comparison of genome-wide DNA methylation in urothelial carcinomas of patients with and without arsenic exposure. Environ Res 128:57-63.

Table 1. Demographic and clinical characteristics of the study population

Characteristic	Value
Maternal Age (years)	31.5 ± 4.9
Pre-pregnancy Maternal BMI	25.6 ± 5.4
Gestational Age (weeks)	39.4 ± 1.6
Infant Birth Weight (kg)	3.4 ± 0.5
Male	183 (53.5%)
Female	159 (46.5%)
Placenta arsenic (µg/kg)^a	0.82(0.47-1.17)
Toenail arsenic (µg/kg)^b	0.05 (0.026-0.073)
Urinary Total arsenic (µg/L)^c	3.76(1.87-5.65)

Values are presented as mean ± SD, n (%), or median (interquartile-range).

^a285 of the 343 mother-infant pairs had available placenta arsenic data

^b257 of the 343 mother-infant pairs had available post-partum toenail arsenic data

^c271 of the 343 mother-infant pairs had available urinary arsenic data at 24-28 weeks gestation

Table 2. CpG loci identified as genome-wide significantly associated (Bonferroni $p \leq 0.05$; $p < 1 \times 10^{-7}$) with placental arsenic levels following cell composition adjustment.

Loci	Coefficient*	FDR q-value	Bon. p-value	Chromosome	Position	Gene
cg01953134	-0.0018	1.73E-07	1.73E-07	6	90348409	LYRM2
cg08640609	-0.0012	5.98E-07	1.75E-06	1	7764642	CAMTA1
cg27642643	-0.0032	5.98E-07	1.79E-06	6	90348450	LYRM2
cg11240327	-0.0042	7.92E-07	3.17E-06	6	90348503	LYRM2
cg10392378	-0.0026	1.32E-06	6.59E-06	6	90348606	LYRM2
cg21621759	-0.0035	7.94E-06	4.76E-05	6	90348302	LYRM2
cg03200341	-0.0023	8.87E-06	6.21E-05	6	90348309	LYRM2
cg21264329	-0.0022	1.68E-05	1.34E-04	6	90348123	LYRM2
cg04843946	-0.0061	1.73E-04	1.56E-03	6	90348009	LYRM2
cg02725437	-0.0025	2.84E-04	2.84E-03	6	90348314	LYRM2
cg25719236	-0.0024	5.60E-04	6.16E-03	6	90348382	LYRM2
cg01777586	-0.0085	3.49E-03	4.19E-02	17	80163174	CCDC57
cg25430506	-0.0038	3.68E-03	4.79E-02	6	90348385	LYRM2

-A complete list of loci may be found in Excel File Table S2

*The difference in DNA methylation β value (0-1 scale) for every one unit ($\mu\text{g/kg}$) increase in placental arsenic.

Table 3. GO-term pathways (Gene Ontology Consortium 2015) significantly associated (Bonferroni < 0.05) with CpG loci with an altered effect estimate due to cellular proportion estimates.

GO ID	Adjusted P-value*	Test Estimate**	GO pathway
9952	2.93E-32	3.267583	Anterior/posterior pattern specification
45944	1.23E-24	1.620397	Positive regulation of transcription from RNA polymerase II promoter
1764	8.85E-20	2.619747	Neuron migration
8045	3.16E-18	4.867233	Motor neuron axon guidance
48704	3.32E-17	3.356625	Embryonic skeletal system morphogenesis
45666	9.08E-17	2.780646	Positive regulation of neuron differentiation
7268	5.81E-16	1.74411	Synaptic transmission
122	6.65E-16	1.555492	Negative regulation of transcription from RNA polymerase II promoter
21884	9.60E-16	7.876803	Forebrain neuron development
7389	2.01E-15	3.83696	Pattern specification process
31018	7.34E-15	3.799679	Endocrine pancreas development
35115	2.13E-14	3.351906	Embryonic forelimb morphogenesis
44267	6.43E-14	0.3915227	Cellular protein metabolic process
21983	7.84E-14	3.640762	Pituitary gland development
1657	9.08E-14	3.162291	Ureteric bud development
21913	1.29E-13	7.137178	Regulation of transcription involved in ventral spinal cord interneuron specification
21912	1.36E-13	8.044255	Regulation of transcription involved in spinal cord motor neuron fate specification
9725	1.36E-13	3.760404	Response to hormone
6351	1.78E-13	1.351092	Transcription, DNA-templated
21797	2.49E-12	5.640727	Forebrain anterior/posterior pattern specification

*Adjusted Bonferroni p-value to account for multiple comparisons

** χ^2_{MH}

Figure Legends

Figure 1. Manhattan plots of epigenome-wide association results for arsenic concentrations within the placenta **(A)**, maternal toenail **(B)**, and maternal urine **(C)**. The horizontal green line represents unadjusted significance threshold ($p = 0.05$) while the blue line represents Bonferroni adjusted significance threshold ($p = 1.0 \times 10^{-7}$).

Figure 2. Volcano plots of the associations between placenta DNA methylation at 344,348 loci and **(A)** placental, **(B)** toenail, and **(C)** total urinary arsenic levels following adjustment for sample cellular composition. The threshold for $p < 0.05$ is indicated by a dashed red line and the threshold for Bonferroni significance ($p = 1.0 \times 10^{-7}$) is indicated by a dashed blue line.

Figure 3. (A) Genomic tracks from the UCSC Genome Browser (University of California Santa Cruz 2015) for the region containing the 11 *LYRM2* loci identified as significantly associated with placental arsenic levels. **(B)** Correlelogram of Pearson correlation test of methylation at these 11 CpG loci, with corresponding heatmap indicating the value of r between -1 and 1..

Figure 4. Scatterplot of the average methylation across the 11 significant *LYRM2* sites and relative expression of *LYRM2* mRNA for 93 individuals with the associated Pearson correlation. Individuals were selected from those within the highest and lowest tertiles of average *LYRM2* methylation.

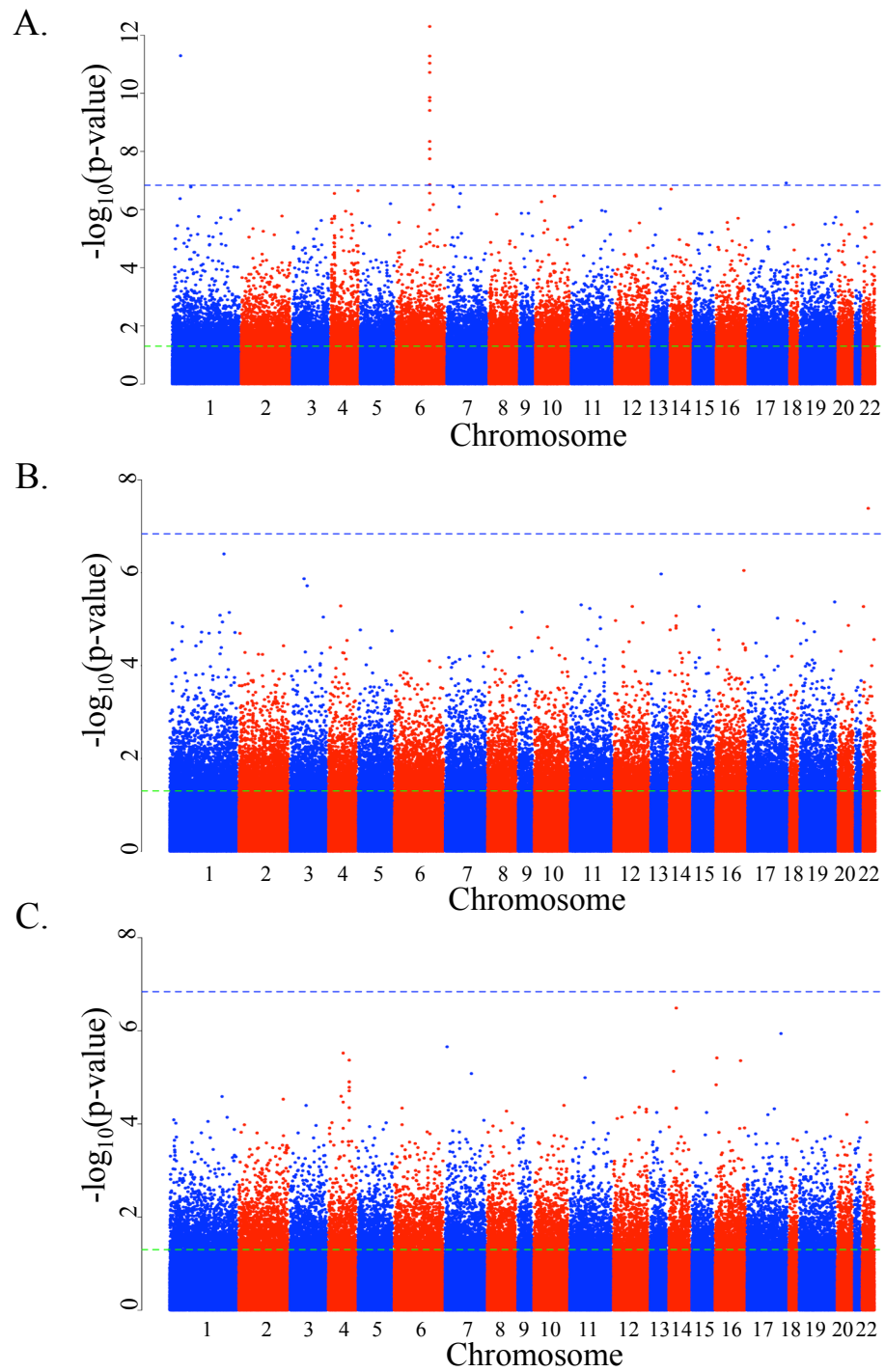


Figure 1

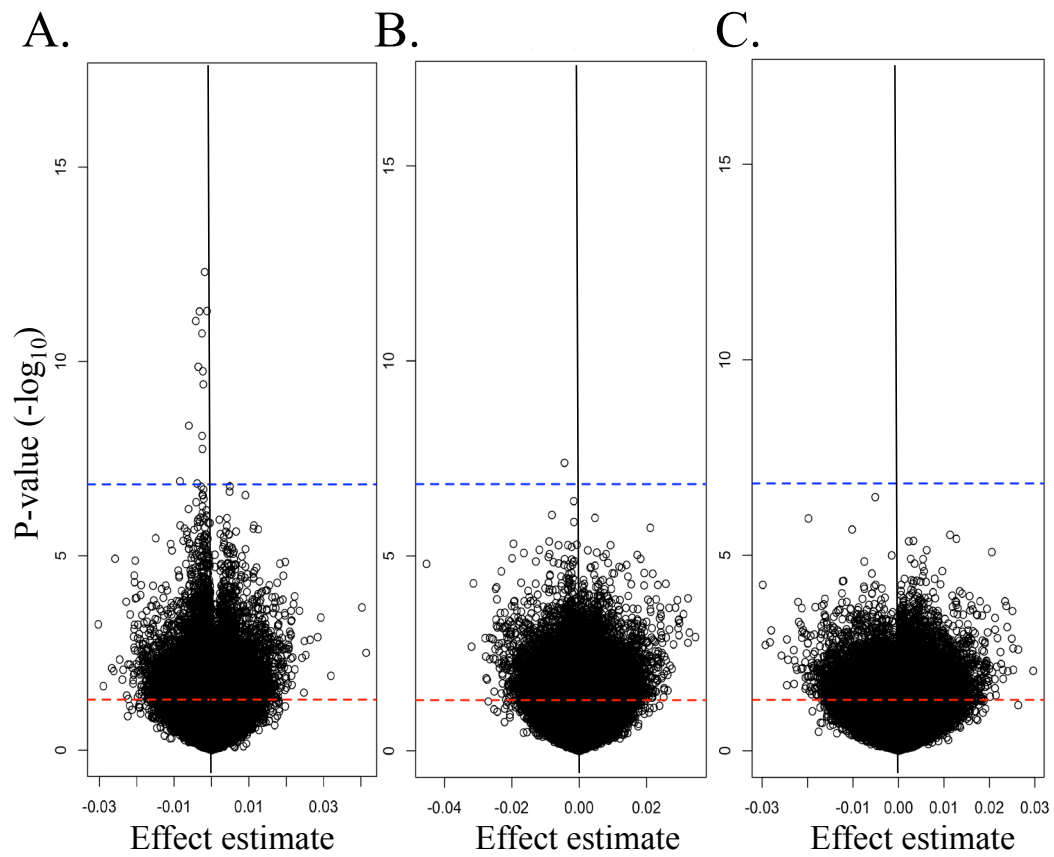


Figure 2

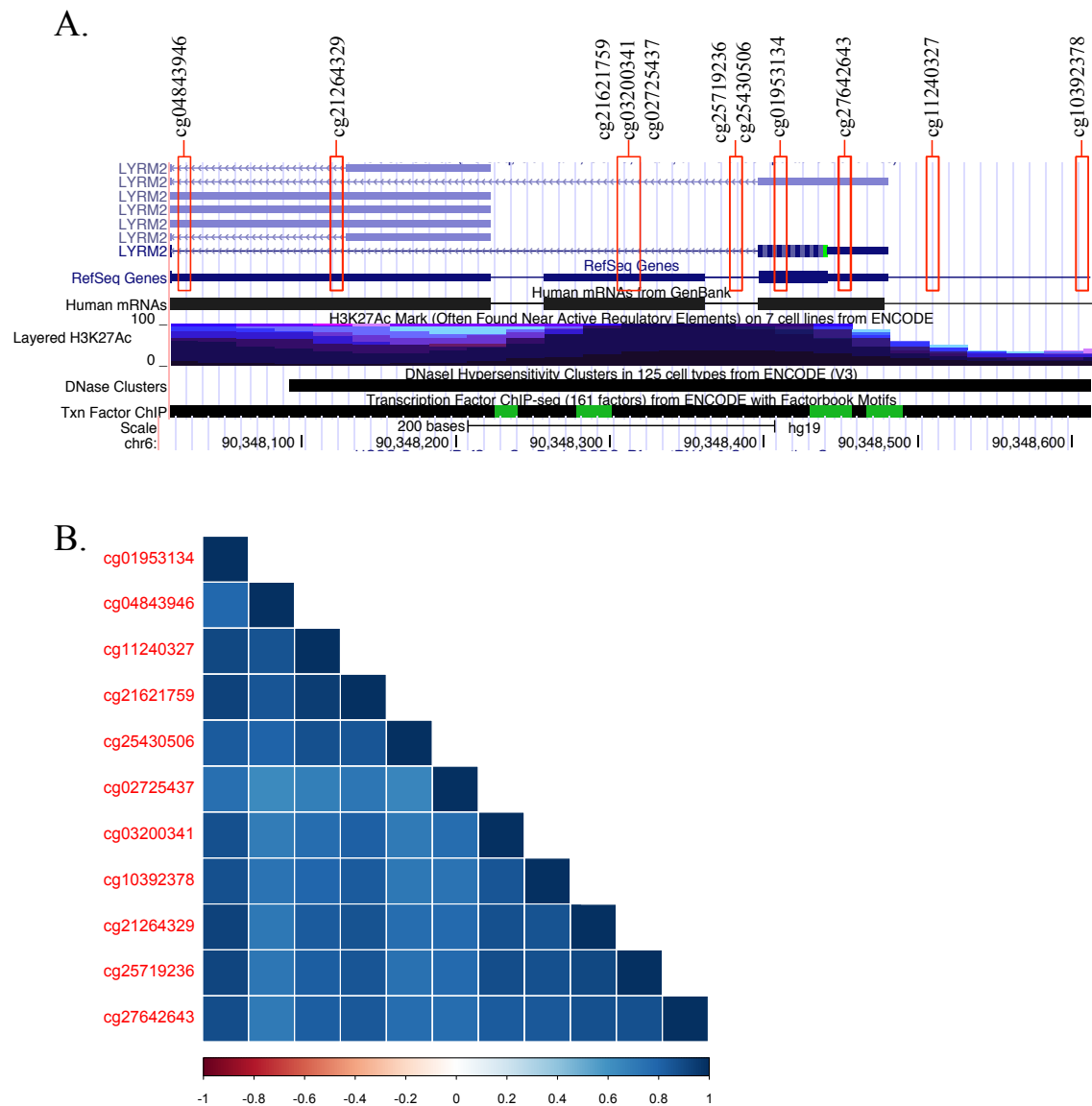


Figure 3

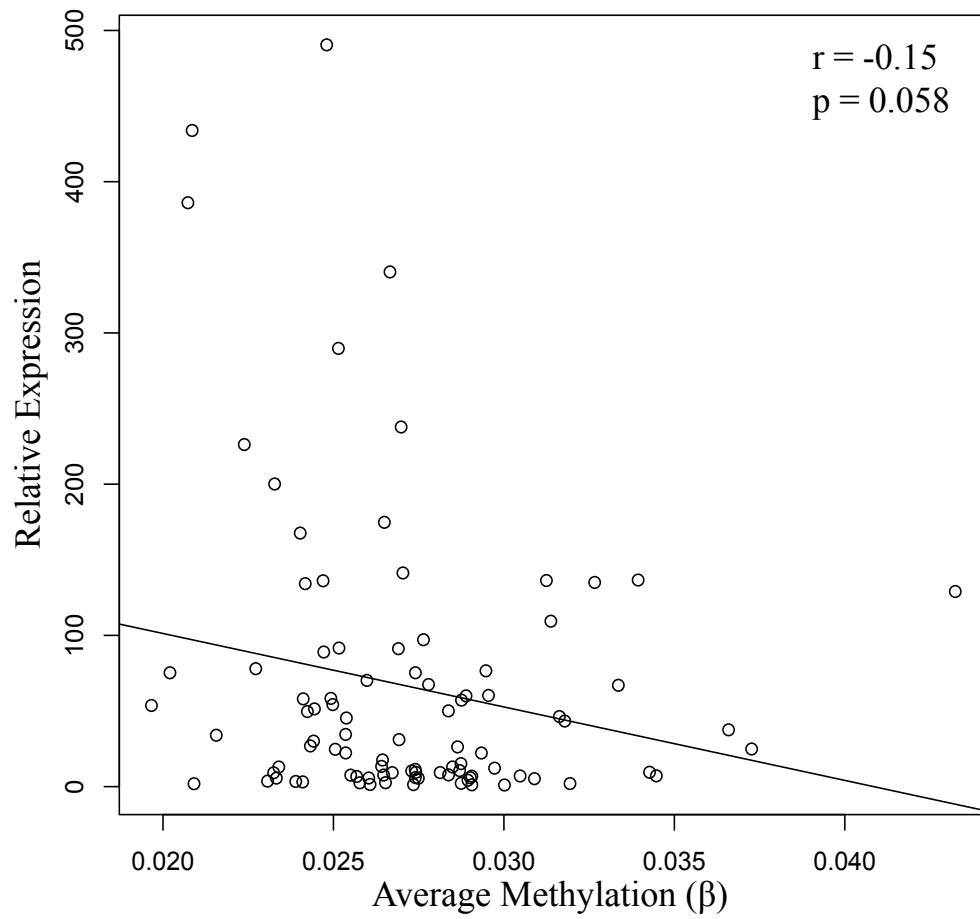


Figure 4